視覚言語モデルにプロンプト最適化と正常画像による補正を適用した異変検知技術

Image Anomaly Detection Method Using Prompt Optimization and Normal-Image-Based Correction for VLMs

河村　直輝 KAWAMURA Naoki　伊藤　聡 ITO Satoshi　瀧本　崇博 TAKIMOTO Takahiro

画像異変検知技術は，産業分野における外観検査や，インフラ点検システム，監視カメラシステムなどに活用される。近年，画像とテキストを事前学習した視覚言語モデル（VLM：Vision-Language Model）の登場で高度な解析が可能になり，異変検知でもVLMの応用研究が注目されている。ただし，その応用には，ユーザーからのプロンプト（テキスト指示）が曖昧な場合に精度が不安定になることや，異変の過剰検知が頻発する問題があった。

そこで東芝は，複数の正常画像を用いてプロンプトを事前に補正することでプロンプトの曖昧さを低減し，更に異変検知の推論時に正常画像を基に補正を行うことで過剰検知を抑制する手法を開発した。開発した手法は公開データセットで評価し，有効性を確認した。

The dissemination of image anomaly detection technologies, which are utilized for various applications, including visual inspection of products in the industrial field, infrastructure monitoring systems, and surveillance camera systems, is driving demand for sophisticated anomaly detection applications with the advent of vision-language models (VLMs), which enable advanced analysis by pre-training jointly on both text and images. However, ambiguous prompts may lead to fluctuations in accuracy or excessive detections of anomalies may frequently occur.

To address these issues, Toshiba Corporation has developed a new image anomaly detection method using multiple normal images to reduce ambiguous prompts by adjusting them in advance, as well as to suppress excessive detections by correcting images when inferring anomaly detection. Experiments using open datasets have confirmed the effectiveness of the new method.

1. まえがき

現代社会では，インフラやプラント設備の保全計画が重要であり，特に国内では，道路，橋，トンネルなどの交通インフラの老朽化に伴い，保守点検作業の必要性が高まっている。一方で，点検作業員の人手不足，点検地点へのアクセスの困難性といった問題を抱えている。そのため，点検作業の省人化のためのAI技術として，画像異変検知技術が注目されている。東芝も，これまで，様々なユースケースに応じた異変検知AIの開発を進めてきた[(1), (2)]。

近年のAI関連分野では，視覚言語モデル（VLM）が登場し，異変検知分野でもVLMの有用性が注目を集めている。VLMは，画像とテキストの両方の情報を結び付けた大規模なデータを用いて事前に学習された基盤モデルであり，従来のAIモデルよりも複雑なタスクに高度なレベルで対応できる。異変検知分野において，VLMは，ユーザーからのプロンプトに基づき，点検画像中の異変箇所・正常箇所を判定する。インフラ点検支援においてもVLMを導入することで，基礎的な異変検知精度の向上のほか，プロンプトを反映することで，現場に応じた柔軟な対応ができるようになることが期待される。

しかし，異変検知において既存のVLMでは，プロンプトが曖昧な場合は検知性能に悪影響が出やすかった。また，点検画像中の複雑な背景構造物を過剰検知しやすかった。

そこで当社は，事前に与えた正常画像を用いて，これらの問題を解決する技術を開発した。ここでの正常画像とは，点検画像と同一の被写体を撮影した，異変のない画像である。開発した手法では，点検画像と正常画像との関係を利用して，ユーザーの入力したプロンプトが曖昧な場合でもVLM向けに自動補正した上で，検知結果における過剰検知を抑制する。これにより，既存技術の異変検知精度を改善し，点検作業の省人化に貢献できる。

ここでは，VLMを用いた異変検知技術と開発した手法に関して述べ，更に，開発した手法の有効性を確認するための実験結果について述べる。

2. VLMを用いた異変検知技術

近年，VLMを活用した異変検知技術（以下，異変検知VLMと略記）が多数提案されている。代表的なものとして，VLMの一種であるCLIP（Contrastive Language-Image Pre-training）を基にしたAIモデルを利用した異変検知技術[(3), (4)]がある。これらの手法では，ユーザーがAIに与える指示で

あるプロンプトを基に，点検画像中の異変領域を検知する。CLIPを基にしたVLMは，画像とテキストのペアデータの大規模セットで事前に学習したモデルであり，画像とテキストを同じ特徴空間に埋め込み学習することを特徴とする。異変検知処理をするには，プロンプトと点検画像をそれぞれ学習済みVLMに入力し，テキストと各画素の特徴ベクトルの類似度に基づき，点検画像の画素ごとの異変度合い（異変スコアマップ）を出力する。プロンプトは，検知対象の異変状態に関するプロンプトと，正常状態に関するプロンプトが各々1文以上用意される。

例えば，鉄道の車載カメラで撮影した路面画像での障害物を異変として検知したい場合，異変プロンプトの例には「a photo of a railroad with obstacle」や，「wide range image of a railroad with obstacle」などが挙げられる。又，正常プロンプトの例には「a photo of a railroad with woods」や，「wide range image of a railroad with building」などが挙げられる。VLMの性能を安定して発揮させるためには，モデルや，タスク，撮影画像に応じて，適切な表現かつ十分な量のプロンプトを用意する必要がある。

一方で，VLMに必要なプロンプトを自動で最適化するための，プロンプト学習[4]という技術がある。この技術では，画像中の被写体に関する最小限のテキスト及びそのテキストを使ったプロンプトの初期表現と，学習用画像を入力とし，プロンプトの埋め込み表現を学習する。これにより，人手によるプロンプトの調整作業を省略できる。

これらの従来技術に関して，主に二つの問題がある。一つ目の問題は，既存のプロンプト学習手法では，画像中の異変の有無を見つけたい被写体に関するテキストや，異変の様相を表すテキストは学習されず，ユーザーの設定した初期値に依存することである。例えば，前述のプロンプト例「a photo of a railroad with obstacle」では，ユーザーは「railroad」（被写体）や「obstacle」（検知対象の異変）に関する最小限のテキスト情報を入力し，プロンプト学習では「a photo of」に対応するテキストの最適なプロンプトを学習する。したがって，「railroad」や「obstacle」といった被写体や検知対象のテキストについては，ユーザーの与える初期値に強く依

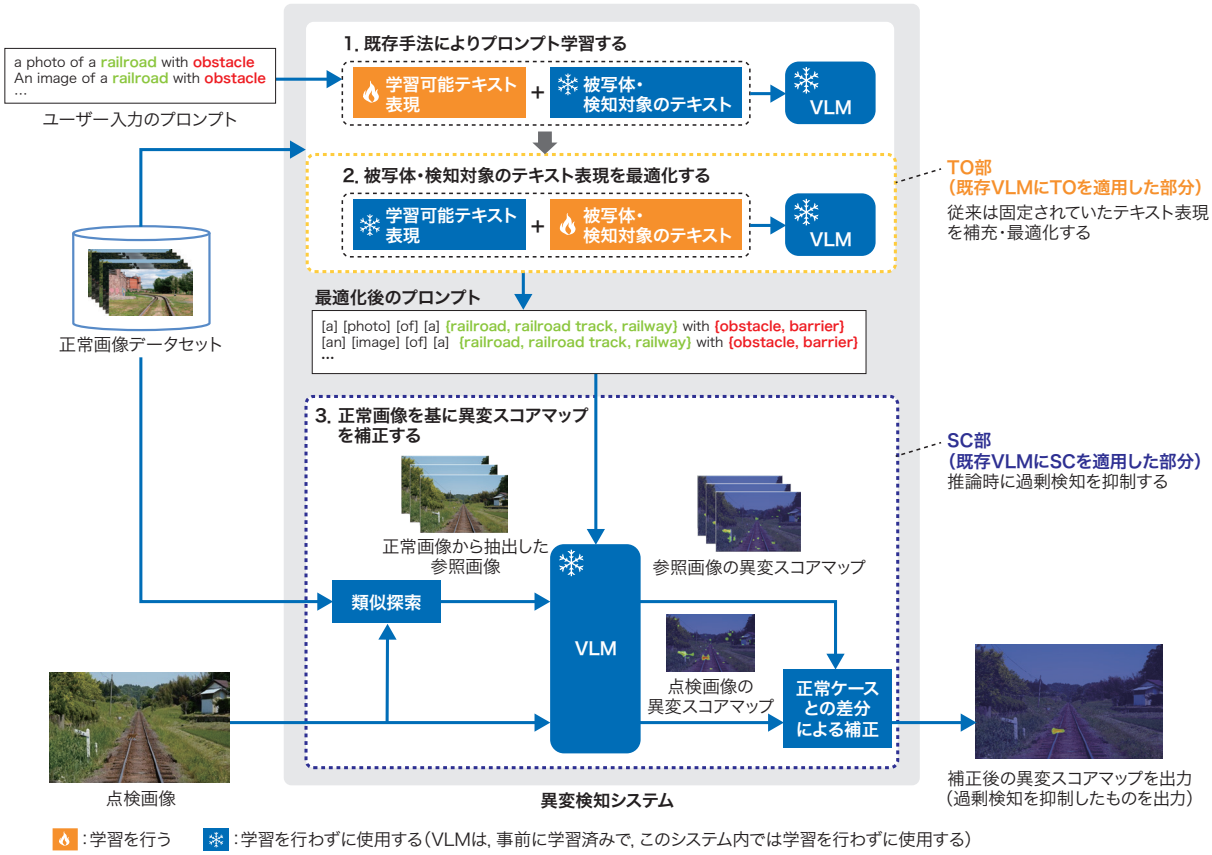

図1. 開発した手法の概要

既存VLMにTO部とSC部を適用した構成にした。TO部でテキスト表現を補充・最適化し，SC部で正常画像を基にした補正で過剰検知を抑制する。

Image anomaly detection method

存していた。このため，ユーザーの入力した被写体や検知対象のテキストの表現が曖昧であったり，不備があったりすると，異変検知精度に悪影響を及ぼすおそれが残されていた。二つ目の問題は，異変検知VLMでは，画像背景領域が複雑化した場合，過剰に検知しやすい傾向が見られることである。画像中の背景構造物が複雑な場合には，適切なプロンプトが十分に用意された場合であっても深刻な過剰検知が発生することがあった。

3. 開発した手法

開発した手法は，異変検知VLMのプロンプト学習における初期値依存性への対処と，異変検知結果での過剰検知の抑制を行うことを課題とし，CLIPを基にした異変検知VLMに対してこれら二つの課題を解決するための枠組みを適用した。

図1に，開発した手法の概要を，2章の鉄道の場合を例に示す。テキスト最適化（TO：Text Optimization）部では，プロンプトに使用するテキストの初期値依存性に対処し，推論時スコア補正（SC：Score Correction）部では，異変検知VLMの過剰検知を抑制する。TO部とSC部のいずれにも，正常画像データセットを事前に与えることと，事前に学習済みのVLMを用いることが，前提である。

3.1 TO部

TO部は，ユーザーの入力したプロンプトにおいて，被写体や検知対象に関するテキスト表現を補充し，プロンプトを最適化する。そのために，既存のプロンプト学習手法と同様に，入力として被写体及び検知対象の異変を表すテキストと，学習用の正常画像を利用する。

まず，被写体と異変を表すテキストの各々について，大規模言語モデル（LLM：Large Language Model）に，言い換え表現を問い合わせる。例えば，LLMによる業務支援向け生成AIサービスに，“railroad（線路）”の10個の言い換え表現を問い合わせる場合，「What are the 10 other expressions of ‘railroad’?」といったプロンプトを入力する。これにより，被写体「railroad」に対して，「railroad track」や「railway」といった10個の言い換え表現が出力される。

次に，正常画像を用いて，各々の言い換え表現を使ったときのプロンプトの妥当性を評価し，正常画像で妥当と判断された言い換え表現を補充したプロンプトを採用する。妥当性の評価には，テキストの言い換え前後で，異変スコアマップの反応度合いを比較する。この反応度合いとして，スコアマップのしきい値以上の領域における平均値を用いる。例えば，正常画像で，「railroad」を使ったプロンプトをVLMに入力するよりも，言い換え表現「railroad track」の方が，スコアマップの反応度合いが良い場合に，「railroad track」を使ったプロンプトを追加で採用する。

これらのTO部の手続きにより，LLMが出力した言い換え表現候補のうち，学習済みVLMにおいて，ユーザー入力のテキスト表現（「railroad」）よりも妥当性の高い言い換え表現（「railroad track」，「railway」）を補充し，異変検知VLMのプロンプトとして与える。

3.2 SC部

SC部は，VLMによる点検画像の異変スコアマップを補正することで，過剰検知を抑制する。この際，TO部と同様に，正常画像を利用する。

最初に，正常画像データセットの中から，点検画像に類似する正常画像を探索し，K枚の参照画像を得る。類似探索では，学習済みVLMの画像特徴空間において，探索手法の一種であるk近傍法を採用して探索した。次に，点検

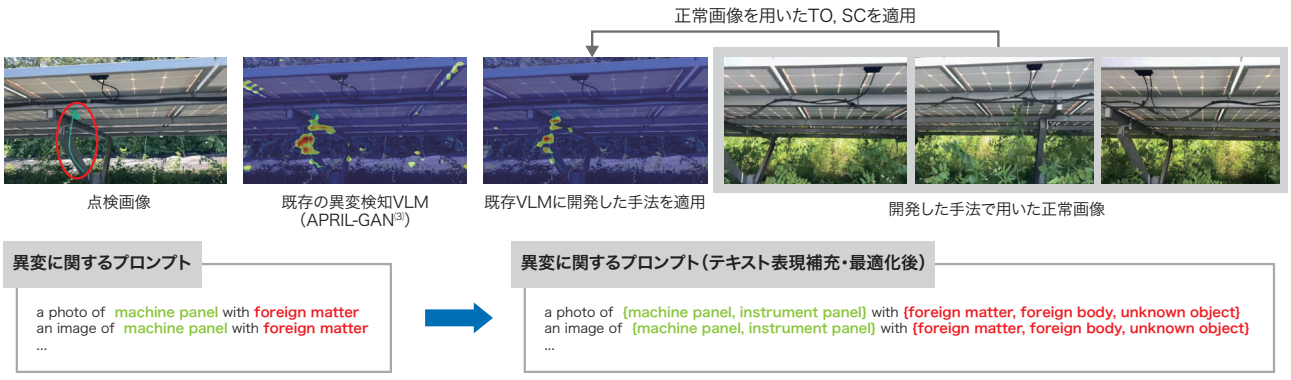

図2. インフラ点検画像における異変検知の例

点検画像の赤線丸部に異変物がある太陽光パネル裏面の例では，開発した手法で3枚の正常画像を用いてテキスト表現を補充し，異常検知を補正したことで，異変検知精度を改善した。

Example of prompt optimization for infrastructure inspection images using normal images

画像及び各々の参照画像と，TO部で出力されたプロンプトをペアとして学習済みVLMに入力し，各々の画像の異変スコアマップを得る。このうち参照画像中の画像は異変を含まないため，参照画像の異変スコアマップは，各画素における過剰検知の元になる部分になると考えられ，この部分の多い／少ないが，過剰検知の発生しやすさの度合いを表す。最後に，点検画像の異変スコアマップと，K枚の参照画像の異変スコアマップの差分を取ることで，点検画像の異変スコアマップを補正する。

これらのSC部の手続きにより，過剰検知を抑制する。そして，TO部とSC部の併用により，ユーザーは最小限のプロンプトの手間でVLMの性能を十分に引き出し，更に高い異変検知性能を得られる。

4. 実験結果

開発した手法の有効性を確認するため，CLIPを基にした異変検知VLMの既存手法を比較対象とし，既存VLMに開発した手法のTO部及びSC部を適用した結果と比較して評価した。比較対象としては，ここ数年で提案された手法であるAPRIL-GAN[3]とPromptAD[4]を選択した。

図2に，APRIL-GANに対して開発した手法を適用した結果を太陽光パネル裏面の例で示す。TO部により，「machine panel」及び「foreign matter」のテキスト表現に対し，新たに「instrument panel」及び「foreign body」，「unknown object」という表現が補充された。更にSC部を組み合わせたことで，過剰検知が抑制され，異変箇所が絞り込まれた。

次に，比較対象の既存手法と開発した手法との定量評価結果を**表1**に示す。評価には，公開データセットであるShanghaiTech Campusデータ[5]を用いた。また，開発した手法は，2種類の既存VLMのそれぞれにTO部，SC部，及びその両方を適用した場合で評価した。異変検知精度の定量評価指標には，正検知率（TPR）95 %時の過剰検知率（FPR）（FPR@95 %_TPR）を採用しており，値が低いほど性能が良い。比較対象の既存手法に対して，既存VLMにTO部だけを適用した場合は，2種類の既存VLMのそれぞれにTO部を適用した結果を平均して，FPRが平均4.6ポイント減少し，SC部だけを適用した場合は平均8.3ポイント減少した。また，TO部とSC部を両方とも適用した場合には，FPRが平均12.4ポイント減少した。これらの結果から，TO部とSC部各々の有効性だけではなく，両方を組み合わせた場合は，更に異変検知性能を改善できることが確認できた。

表1. 公開データセットでの定量評価結果

Quantitative evaluation results using open datasets

評価に用いた既存VLM	適用した開発した手法		FPR@95 % _TPR（%）	比較対象（既存手法）からの減少分（改善分）（ポイント）
	TO	SC		
APRIL-GAN			32.1	比較対象
	✓		25.7	6.4
		✓	20.7	11.4
	✓	✓	**16.1**	**16.0**
PromptAD			30.6	比較対象
	✓		27.8	2.8
		✓	25.4	5.2
	✓	✓	**21.9**	**8.7**

5. あとがき

異変検知VLMの性能改善のための枠組みとなる技術を開発した。開発した手法は，点検画像と正常画像との関係に基づき，ユーザーの曖昧なプロンプトを自動で補正し，かつ異変検知VLMの推論時に過剰検知を抑制することで，異変検知精度を改善する。公開データセットを用いて，既存手法を比較対象とし，2種類の既存VLMに対して開発した手法のTO部，SC部，及び両方を適用して，それぞれの効果を評価した実験の結果，有効性を確認した。

開発した手法は，車載カメラや，定点カメラ，ドローンカメラなどを用いた，交通インフラやプラント設備の点検支援に応用できる。今後は，様々な設備点検シーンでの実証実験を進める。同時に，商用化を想定したモデルのブラッシュアップを図り，実用に向けて研究開発を進めていく。

文　献

(1) 東芝．“インフラ点検向けに数枚の正常画像から異常個所を世界最高精度で検出するAIを開発”．研究開発ニュース．<https://www.global.toshiba/jp/technology/corporate/rdc/rd/topics/22/2205-01.html>,（参照 2026-01-05）.

(2) 東芝，東芝デジタルソリューションズ．“高速道路の保全と長期稼働に貢献する路面変状検知AIを開発し、重大事故につながる危険のある路面の穴のリアルタイム検知を実証”．研究開発ニュース．〈https://www.global.toshiba/jp/technology/corporate/rdc/rd/topics/23/2309-01.html>,（参照 2026-01-05）.

(3) Chen, X. et al. "A Zero-/Few-Shot Anomaly Classification and Segmentation Method for CVPR 2023 VAND Workshop Challenge Tracks 1&2: 1st Place on Zero-shot AD and 4th Place on Few-shot AD". The IEEE / CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2023. Vancouver, Canada, 2023-06, IEEE Computer Society, 2023, arXiv:2305.17382.

(4) Li, X. et al. "PromptAD: Learning Prompts with only Normal Samples for Few-Shot Anomaly Detection". The IEEE / CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2024. Seattle, WA, 2024-06, IEEE Computer Society, 2024, p.16838–16848.

(5) Liu, W. et al. "Future Frame Prediction for Anomaly Detection – A New Baseline". The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2018. Salt Lake City, UT, 2018-06, IEEE Computer Society, 2018, p.6536–6545.

河村　直輝　KAWAMURA Naoki, D.Eng.
総合研究所 AIデジタルR&Dセンター アナリティクスAI研究部
博士(工学)
Analytics AI R&D Dept.

伊藤　　聡　ITO Satoshi
総合研究所 AIデジタルR&Dセンター コラボレイティブAI研究部
Collaborative AI R&D Dept.

瀧本　崇博　TAKIMOTO Takahiro
技術企画部 技術戦略企画室
Strategic Technology Planning Office